American Journal of Medical Genetics (Neuropsychiatric Genetics) 67:393–400 (1996)

Identification of Symptomatologic Patterns Common to Major Psychoses: Proposal for a Phenotype Definition

Alessandro Serretti, Fabio Macciardi, and Enrico Smeraldi

Department of Neuroscience, School of Medicine, Istituto Scientifico H. San Raffaele University of Milan, Milan, Italy

Our study was designed to identify the underlying symptomatologic structure common to major psychoses as a preliminary step for a phenotype definition. We investigated 1,004 inpatients affected by mood disorders or the schizophrenia spectrum (DSM-III-R) using the OPCRIT checklist (operational criteria checklist for psychotic illness). Symptomatologic structure was extracted by factor-analytic techniques and factor scores were first obtained on 500 subjects. A CFA (confirmatory factor analysis) was then conducted on the remaining 504 subjects to evaluate fitness of the model. We identified four factors: excitement, depression, disorganization, and delusion. These factors accounted for 54.6% of the total variance of the OPCRIT checklist symptomatologic subset of 38 items. CFA indices showed a good fit for the model. We identified symptomatologic structures common to major psychoses. The factors identified were confirmed in an independent sample. Two of these symptomatologic structures are partially overlapping with categorical diagnoses (excitement and depression), and two constitute independent psychopathologic traits (delusion and disorganization). The use of "factor-derived scores" in genetic research may add a dimensional definition to the diagnostic subdivision of major psychoses. © 1996 Wiley-Liss, Inc.

KEY WORDS: factor analysis, phenotype, schizophrenia, depression

INTRODUCTION

The definition of phenotype is one of the most important problems in psychiatric genetic research [McGuffin et al., 1991; Tsuang and Faraone, 1990; Farmer et al., 1994; Rutter, 1994]. The absence of any identified biologic marker forces the researcher to rely on the observed psychopathology of the disorder and, when possible, on external validations such as the course of the disorder or the family history [Faraone and Tsuang, 1994]. Psychopathologic symptoms tend to cooccur, forming clusters that are the basis for the definition of psychiatric syndromes. Although syndromes are the most popular way to define phenotypes in psychiatric genetic research, it has been argued that the intraclass variability observed within the same psychiatric syndrome could constitute a potential bias for detecting underlying genetic factors [McGuffin et al., 1991; Farmer et al., 1994]. Additionally, a less heterogeneous phenotype description is required to increase the power of association or linkage studies [McGuffin et al., 1991; Farmer et al., 1994].

To improve the phenotype definition, the existence of latent roots of psychopathology underlying psychiatric syndromes has been hypothesized. These latent roots, more general than single symptoms, but less general than psychiatric syndromes, would produce the observed symptomatology. The search for latent structures can be formalized through mathematical modeling [Lieberman, 1995; Farmer et al., 1994; Tsuang, 1993]. For schizophrenia, a number of reports proposed and tested models of 1–5 underlying latent roots [Andreasen and Olsen, 1982; Crow, 1980; Andreasen et al., 1995; Liddle and Barnes, 1990; Peralta et al., 1994; Von Knorring et al., 1995; Maziade et al., 1995]. The one-factor model binds all schizophrenia symptoms into one factor with positive and negative symptoms at the extremes [Andreasen and Olsen, 1982]. The classification of schizophrenia into types I and II is an example of the two-factor model [Crow, 1980], also confirmed by a factor-analytic study [Mortimer et al., 1990], where "positive" and "negative" factors included all symptoms. A three-factor model composed of "negative" (blunted affect, alogia, avolition, and anhedonia), "positive" (hallucinations and delusions), and "disorganized" (positive and negative thought disorders, bizarre

Received for publication October 19, 1995; revision received February 26, 1996.

Address reprint requests to Fabio Macciardi, M.D., Ph.D., Department of Neuroscience, Istituto Scientifico H. San Raffaele, University of Milan-School of Medicine, Via L. Prinetti, 29, 29127 Milan, Italy.

behavior, and inappropriate affect) symptoms was proposed by the majority of researchers, and appeared to be quite stable among different samples [Bilder et al., 1985; Liddle and Barnes, 1990; Andreasen et al., 1995; Maziade et al., 1995]. A few studies proposed a fourth "relational" factor (intimacy and closeness, and relationships) that proved to be highly correlated with the "negative" pattern [Peralta et al., 1994].

For mood disorders, far fewer reports have been produced. For bipolar disorder, to our knowledge, only a three-factor model was detected, composed of the same "negative," "positive," and "disorganized" factors found in schizophrenia [Maziade et al., 1995]. For major depression as well, models composed of three distinct factors have been proposed, mainly subdividing depressive and anxiety symptomatology, with a third minor factor regarding behavior [Kendler et al., 1987; Maes et al., 1994; Parker et al., 1993]. The same three factors detected in schizophrenic and bipolar samples were also found in heterogeneous psychotic populations [Minas et al., 1992; Klimidis et al., 1993].

The great majority of these studies were performed using factor analyses, but other exploratory techniques have sometimes been applied, such as cluster analysis in Minas et al. [1992] or in Maes et al. [1994]. The latent-roots approach constitutes a powerful method for phenotype definition in genetic research, since it greatly increases the power of detecting liability effects in a basic psychopathologic area [Rutter, 1994; Cloninger, 1994; Lieberman, 1995] which could be common to distinct syndromes. However, only a few studies applied a latent-roots perspective to phenotype definition in genetic research [Kendler et al., 1987; Eaves et al., 1993]. Moreover, most attempts to identify latent structures were performed on diagnostically homogenous samples. While this is largely justified by an interest in finding subsyndrome structures, it does not allow detection of liabilities common to two or more psychiatric syndromes.

The process of identification of latent structures with factor analysis may be biased, since the appropriateness of the factorization and its stability are the major concerns. To test that appropriateness of the factorization, the use of confirmatory factor analysis (CFA) on a different sample to substantiate the results of a previous exploratory factor analysis (EFA) is now commonly accepted [Reyment and Joreskog, 1993; Bollen, 1989; Joreskog and Sorbom, 1989]. Regarding stability, it has been proven that latent class estimators are asymptotically correct, and thus that only large samples allow meaningful estimates [Joreskog and Sorbom, 1989; Bollen, 1989].

The purpose of our study was to detect latent structures underlying major psychoses. We performed an EFA on a sample of patients affected by schizophrenic and mood disorders. Then we confirmed our results with CFA on an independent sample.

MATERIALS AND METHODS

Subjects

Subjects in this study consisted of 1,004 consecutive patients admitted to the Department of Neuropsychiatry at the Institute H. San Raffaele (DSNP-HSR), units for schizophrenic and depressive disorder. The demographic and clinical features of subjects are summarized in Table I.

All patients admitted to DSNP-HSR received an extensive evaluation that included the Schedule for Affective Disorders and Schizophrenia (SADS) [Endicott and Spitzer, 1978] and/or the Diagnostic Interview Schedule (DIS) [Robins et al., 1981], as well as the OPCRIT checklist (operational criteria checklist for psychotic illness) [McGuffin et al., 1991], and were diagnosed according to DSM-III-R criteria [American Psychiatric Association, 1987]. Consensus diagnosis was performed by two independent psychiatrists, one of whom was a senior psychiatrist. Diagnostic criteria for inclusion in the study were represented by clinical diagnoses of major psychoses: schizophrenia, including all subtypes; depressive disorder, not including dysthymia; bipolar disorder, not including cyclothymia; delusional disorder; schizophreniform disorder; and psychotic disorder not otherwise specified. Patients also underwent a detailed medical examination. All patients meeting criteria for a medical or neurological disorder markedly influencing psychiatric status (e.g., hypothyroidism mimicking a depressive state), and patients presenting DSM-III-R Axis I comorbidity (e.g., organic

TABLE I. Demographic and Clinical Data

	Group A		Group B	
Age	41.61 (±14.15)		42.49 (±13.84)	
Age at onset	28.59 (±12.73)		30.24 (±12.45)	
Male/female	223/277	44.6%/55.4%	341/163	67.7%/32.3%
Unmarried	269	53.81%	239	47.42%
Unemployed	248	49.58%	242	48.02%
Diagnoses				
Bipolar disorder	145	29.00%	162	32.14%
Delusional disorder	39	7.80%	42	8.33%
Psychotic disorder NOS	49	9.80%	49	9.72%
Schizophrenia	159	31.80%	147	29.17%
Schizophreniform	2	0.40%	5	0.99%
Unipolar depression	106	21.20%	99	19.64%
Total	500	100.00%	504	100.00%

mental disorder, psychoactive substance abuse or dependence, or obsessive compulsive disorder) were excluded from the study.

The sample was then randomly divided into two subsamples of 500 (group A) and 504 (group B) patients. Group A was used for an exploratory factor analysis, and group B for a confirmatory factor analysis. The two groups did not differ regarding clinical or sociodemographic variables.

Rating Procedure

Symptoms were assessed using the Operational Criteria Checklist for Psychotic Illness version 3.3 (OPCRIT 3.3) [McGuffin et al., 1991]. This is a polydiagnostic 90-item checklist specifically designed for genetic research in the context of the Network on the Molecular Neurobiology of Mental Illness, a collaborative study of the European Science Foundation. The checklist was compiled by trained psychiatrists relying on unstructured interviews and clinical records. As suggested by Farmer et al. [1994], we applied a lifetime perspective for scoring symptoms.

The 90 OPCRIT variables were reduced to 38 for input of the factoring process.

This reduction was conducted following the criteria of: a) Exclusion of variables not directly related to phenomenology (e.g., source of rating, time frame, etc.); b) Exclusion of variables with variance near 0, i.e., scoring 0 for almost all subjects (e.g., lifetime diagnosis of cannabis abuse, cannabis abuse/dependence with psychopathology); and c) Exclusion of variables showing high colinearity with other variables, i.e., variables that slightly differed from others: these variables were originally created for purposes of internal consistency, but they were not informative for our aim and often resulted in having exactly the same score in our sample (e.g., persecutory/jealous feelings and hallucinations vs. abusive/accusatory/persecutory voices).

Statistical Analysis

Differences between continuous variables were assessed using a t-test. Frequencies were compared by chi-square test with Yates correction, or with Fisher's exact test when required.

Factor Analysis

Aiming to identify the best factorial model for our sample, and lacking, to our knowledge, a well-established methodological procedure to deal with heterogeneous populations, we decided to apply a two-stage process: first, we used an exploratory factor analysis (EFA) to identify latent structures in group A, and then we tested the model with a confirmatory factor analysis (CFA) to evaluate the fit with our data in group B.

EFA. EFA was performed using the principal component analysis method on the subset of 38 items of the OPCRIT checklist. Because of the ordinal or dichotomous nature of our data, we used a polychoric correlation matrix, since it has been demonstrated that this particular matrix gives more stable parameter estimates when using ordinal or dichotomous variables [Joreskog and Sorbom, 1989]. Direct quartimin oblique rotation was allowed, since latent psychological structures are usually expected to be interrelated [Peralta et al., 1994; Liddle and Barnes, 1990].

The number of factors to be included was defined following the criteria of 1) eigenvalues greater than average; in our case, for a polychoric matrix the eigenvalue average is 1; 2) evaluation of the shape of the eigenvalue function, where a strong change on the slope of this function may indicate that any other factor following the abrupt slope variation should be disregarded (all other factors may represent only distortion of the overall pattern, being mostly only "error factors").

The definition of number of factors in factor analysis is of crucial importance, especially because this is the main difference between EFA and CFA. In EFA, all variance of original data is explained by factors, which at the beginning of the analysis are equal in number to the original variables. In CFA the number of factors is specified a priori and the data variance not explained by those factors is defined as random variance, and is unique to each variable. This unexplained variance may be due to measurement errors or to the uniqueness of the variable, and is considered not correlated with the other variables. Thus, an underestimation of the number of factors leads to the consideration of some latent structures as errors, while overestimating their number leads to the classification of part of the error as factors.

Then the estimated factor scores for all factors were assigned to each patient. The estimated factor scores were obtained through the following regression equation:

$$\xi = \Phi\Lambda'_x \Sigma^{-1}_x$$

where ξ is the vector of estimated factor scores, Φ is the covariance matrix of factors, Λ'_x is the transpose of the factor loadings matrix, Σ^{-1} is the inverse of the covariance matrix of data, and x is the data vector. These factor scores are standardized, so that each factor score distribution has a mean of 0 and an SD of 1. The estimated factor scores are then considered as new variables, and the distribution across diagnoses is calculated: i.e., following EFA, we performed an extended EFA on the results obtained so far, evaluating the cross correlation between the factors identified by the exploratory analysis and the diagnoses obtained by clinical investigation.

CFA. The second step consisted in CFA performed on group B. CFA is a confirmatory technique that requires specification of a model of latent structures. The differences between CFA and EFA consist primarily in the possibility to define as fixed or free the elements of the Λ, Φ, and Θ_δ (covariance matrix of errors) matrices. The elements set free are estimated according to the maximum likelihood procedure, while the elements set fixed are fixed to zero. This means that for the Λ matrix the variables are constrained to load only on one factor, allowing a clearer interpretation of the factors. The matrix resulting as output from the EFA (Λ) is introduced as input for the CFA, and the model tests whether this solution is the best possible, given the observed data. The fitting of the model is evaluated through the following indices:

1. Chi-square. This is derived from the maximum likelihood fit function, and is distributed asymptot-

ically as a chi-square distribution: a probability value can then be derived from it. This is the probability of rejecting the null hypothesis H0: the model fits the data. Hence, a significant chi-square is an indication of bad fit of the model, but this indicator is very sensitive to sample size and even more to multivariate normality [Joreskog and Sorbom, 1989]. Since our data are represented mostly by ordinal or even dichotomous variables, our sample may show a certain degree of departure from multivariate normality. This indicator in our sample may then be regarded only as a goodness-of-fit indicator for comparisons between models.

2. Goodness-of-fit index (GFI). This is a function ($GFI = 1 - [tr(\Sigma^{-1} S - 1)^2]/[tr(\Sigma^{-1} S)^2]$, where S is the covariance matrix of the observed data) derived from the difference of the variance of the estimated model and the variance of the observed data, thus representing an index of how much variance of the data is predicted by the model. The number of subjects is not included in this index, which consequently should be independent from sample size. Value ranges from 0–1, and values >0.90 are index of good fit [Bollen, 1989].

3. Adjusted GFI. Another bias affecting the model evaluation is overfit, i.e., a condition where the number of explanatory variables is too high relative to sample size. Since GFI could be artificially inflated by an excessive number of independent estimated parameters, AGFI adjusts GFI for degrees of freedom of the model relative to number of variables. The range is the same as in GFI (0–1), and a good-fitting value is AGFI >0.80 [Bollen, 1989].

4. Root mean square residual. This is derived from the difference between observed and predicted elements of the covariance matrix. A low value is considered an index of good fit, but this value is considered affected by sample size.

Other useful indicators are t-values and modification indices. The t-values are the ratio of the parameter estimate to its standard error, and significant t-value indicates that the parameter should most likely not be set to zero. The modification index is derived by differentiating the likelihood function at the unrestricted parameter vector, and can be interpreted as the chi-square change obtained as the fixed parameter is set free.

All statistical procedures were performed using the BMDP [Dixon, 1990] and LISREL 8 computer packages [Joreskog and Sorbom, 1993].

RESULTS

We first performed EFA with the specification explained in the previous section, and a four-factor model was chosen. There were two reasons for this choice. As we have already pointed out, it is recommended to retain all factors with eigenvalues >1, and in our sample this procedure selected four factors. Moreover, after the fourth factor a major change in the slope of the eigenvalues function occurred [Joreksog and Sorbom, 1989]. In our sample this means that each factor from 5th–38th explain <1% of the variance, i.e., less than a single item variance. Thus, a four-factor solution appears to be the best-fitting model.

We also tried different specifications of number and rotation of factors, but we substantially obtained the same four factors. The first four factors accounted for 54.6% of the total variance. Factors resulting from EFA are shown in Table II.

The factor-loading distribution in Table II presents a clear partition of symptoms, with only one item loading on more than one factor ("early morning waking" on "depression" and "excitement" factors). We defined the first factor, "excitement," the second, "depression," then "disorganization," and the last, "delusion." Since we allowed for correlation among factors, "excitement" and "depression" showed a slight positive correlation, as did "disorganization" and "delusion." These results are not unexpected, since schizophrenic subjects often present a lifetime combination of both disorganized and delusional symptoms, but bipolar patients also have, by definition, a history of depressive and "excitement" symptoms.

Extended EFA

While EFA produced a factor score for each patient, Table III shows results of the extension analysis performed on individual diagnoses. For each DSM-III-R diagnosis we present the mean factor score. For the most part, bipolar patients reported high scores on the "excitement" factor. On the "depression" factor, high scores were reported by almost all depressed, all delusional depressed, and about 75% of bipolars, but on this factor the distribution of bipolar subjects shows a certain degree of bimodality. This bimodality is due to those bipolar subjects with severe episodes of mania but mild episodes of depression. On the factor "disorganization," mainly schizophrenia (including schizophreniform) and psychotic disorder not otherwise specified (NOS) patients scored high; a small group of bipolar patients also scored high, while none of the depressive subjects did. On the "delusion" factor, schizophrenics (including schizophreniforms), delusionals, some depressives with psychotic features, and some psychotic NOS patients reported high scores.

CFA

To evaluate the fit of the factor structure obtained by EFA, we used the factor-loading matrix of Table II with all loadings <0.300 fixed to 0 because they were not significantly different from 0. CFA performed on independent sample B showed a substantial good fit of the model ($\chi^2 = 3472.88$; $df = 659$; $P < 0.0001$; GFI = 0.829; AGFI = 0.803; RMSR = 0.081). The significant chi-square value, as stated in the previous section, must be considered cautiously, as it can be inflated by departure of the sample from the multivariate normality. GFI is an indication of good fit. The significantly high AGFI is also a support of good fit. The small difference between GFI and AGFI indicates that our model uses a low number of parameters, and a parsimonious model is always preferable to a more parametrized one. The t-values of each loading coefficient ranged from 11.37–51.52 and this means that all variables included in the model should remain included.

The modification indices are all quite low. The higher indices indicate that for each unfixed parameter, the chi-square value would decrease to <70, and the consequent correlation value would never be higher than

TABLE II. Best-Fit Result of EFA for 38 Items of OPCRIT Checklist Over 504 Subjects*

Variables	Excitement	Depression	Disorganization	Delusion
Excessive activity	**0.875**	−0.040	−0.064	−0.088
Reduced need for sleep	**0.875**	−0.007	−0.077	−0.098
Pressured speech	**0.864**	−0.102	−0.075	−0.077
Elevated mood	**0.863**	0.003	−0.064	−0.108
Thoughts racing	**0.859**	−0.038	−0.093	−0.034
Increased sociability	**0.799**	−0.023	−0.045	−0.079
Increased self-esteem	**0.775**	−0.05	−0.113	0.046
Irritable mood	**0.746**	−0.024	−0.005	0.013
Distractibility	**0.730**	0.019	0.299	−0.088
Agitated activity	**0.606**	0.088	0.187	0.092
Dysphoria	**0.576**	0.232	0.017	0.061
Grandiose delusions	**0.569**	−0.056	−0.013	0.279
Reckless activity	**0.554**	0.113	0.077	−0.061
Loss of pleasure	−0.079	**0.842**	0.011	0.033
Loss of energy/tiredness	−0.108	**0.786**	0.085	0.006
Diminished libido	−0.081	**0.771**	−0.019	−0.029
Excessive self-reproach	0.032	**0.745**	−0.175	−0.044
Slowed activity	−0.014	**0.710**	0.141	0.013
Poor appetite	0.006	**0.701**	−0.167	−0.056
Poor concentration	0.119	**0.646**	0.221	0.015
Suicidal ideation	−0.005	**0.603**	−0.123	−0.004
Weight loss	0.040	**0.584**	−0.117	−0.005
Diurnal variation	0.189	**0.557**	−0.271	−0.027
Early morning waking	0.315	**0.518**	−0.196	−0.083
Speech difficult to understand	−0.015	0.027	**0.720**	−0.091
Incoherent	0.154	−0.078	**0.719**	−0.050
Positive formal thought disorder	0.068	−0.071	**0.610**	0.195
Inappropriate affect	0.047	−0.081	**0.600**	−0.071
Bizarre behavior	0.093	−0.118	**0.584**	0.139
Blunted affect	−0.196	0.114	**0.570**	0.147
Negative formal thought disorder	−0.197	−0.021	**0.512**	0.112
Deterioration from premorbid level of function	−0.252	−0.144	**0.498**	0.243
Persecutory delusions	0.023	−0.034	−0.005	**0.766**
Persecutory/jealous delusions, and hallucinations	−0.063	−0.075	0.085	**0.701**
Well-organized delusions	0.021	0.029	−0.272	**0.697**
Widespread delusions	−0.064	0.068	0.102	**0.691**
Delusions and hallucinations last for 1 week	−0.004	−0.021	0.167	**0.684**
Delusions of influence	0.015	−0.052	0.106	**0.673**
Variance explained	**9.71**	**5.54**	**3.45**	**2.04**
Cumulative variance explained (%)	**25.58**	**40.17**	**49.25**	**54.63**

* For each factor we reported the factor-loading values. Higher values indicate strong correlation between symptom and factor. Bold values show composition of each factor.

0.18. Neither variables should be removed from the model, nor should a substantial increase in fit be obtained if any variable is allowed to load on more than one factor.

DISCUSSION

The aim of this study was to define latent roots of psychopathology, and indeed we identified symptomatologic structures underlying categorical diagnoses. We also confirmed these findings in an independent sample. Previous studies showed substantial agreement that three factors underlie schizophrenic disorder: positive, negative, and disorganized symptoms [Andreasen and Olsen, 1982; Crow, 1980; Andreasen et al., 1995; Liddle and Barnes, 1990; Peralta et al., 1994; Von Knorring and Lindstrom, 1995; Maziade et al., 1995]. A minority of researchers found a different number of factors [Peralta et al., 1994; Von Knorring and Lindstrom, 1995]. Although most of these studies were performed on homogeneous schizophrenic samples, they identified factors very close to ours. Our "delusion" is identical to the "positive" factor, and our "disorganized" contains mainly items from the "disorganized" factor and one ("blunted affect") from the "negative" factor. However, our sample also included mood disorders, and thus we extracted other factors that relate to the mood area, i.e., "excitement" and "depression."

As stated in Materials and Methods we allowed for correlation between factors, and thus "disorganization" and "delusion" factors proved to be positively correlated with each other, as well as "excitement" and "depression" factors. A possible explanation for these correla-

TABLE III. EFA: Factor-Score Distribution Among Clinical Diagnoses

	Excitement		Depression		Disorganization		Delusion		
DSM III-R diagnoses	Mean	SD	Mean	SD	Mean	SD	Mean	SD	n
Bipolar	1.192	0.880	0.335	1.078	−0.382	0.614	−0.474	0.909	307
Unipolar	−0.667	0.420	0.971	0.768	−0.689	0.445	−0.598	0.860	153
Unipolar (with psychotic features)	0.461	0.643	1.107	0.554	−0.829	0.381	0.055	0.740	52
Schizophreniform	−0.361	0.334	−0.379	0.413	0.534	0.983	0.719	0.642	7
Schizophrenia	−0.500	0.347	−0.548	0.571	0.933	0.967	0.716	0.788	306
Delusional	−0.633	0.241	−0.622	0.661	−0.634	0.60	0.572	0.694	81
Psychotic NOS	−0.360	0.574	−0.412	0.697	0.300	0.935	0.019	0.755	98
All groups	0.024	1.006	0.016	1.004	0.025	1.013	0.052	0.994	1,004

tions is the lifetime perspective of scoring: bipolar patients will then show both manic and depressive symptoms, and schizophrenic patients are also quite likely to present clinical pictures composed of positive and disorganized features. Since all other correlations were negative, we can associate these two aggregations ("excitement" and "depression," "disorganization" and "delusion") with the two main clinical diagnoses of mood and schizophrenic disorders.

To our knowledge, our study employs the largest number of subjects as compared with published reports, and this allowed us to split the original sample into two groups (A and B). In this way we were able to create a model with sample A, and to confirm it, using independent sample B. We obtained good values-of-fit indices, thus suggesting that no further improvement of fit could be obtained by marginal model modifications. The AGFI value is only slightly beyond the threshold of good fit, but the difference with GFI is very small. AGFI is a fitting index adjusted for number of parameters in the model, with more parameters corresponding to a decrease of the fitting index, and this is justified by the risk that an overparametrized model would fit the data anyway. On the contrary, our model is highly conservative, given that only four factors were included in the model to explain 38 variables.

"Excitement" was the first factor found, and it explains the greater amount of variance in our sample. The high variance means that the variables loading on it are highly interrelated. Also, a simple visual inspection of the correlation matrix reveals a clear pattern of association among symptoms of the manic state. This is not surprising, as the set of symptoms of the manic state has been proved to have a very high degree of construct validity [Young et al., 1983]. This factor represents a latent structure very specific to bipolar disorder, because bipolar subjects scored very high on it, and all other groups scored very low.

The "depression" factor was composed mainly of typical major "depression" symptoms. Previous analyses of depressive symptomatology yielded a depressive factor composed essentially of the DSM-III-R criteria for major depression [American Psychiatric Association, 1987]. These estimates seemed to be very stable among different samples [Eaton et al., 1989; Parker et al., 1993], except for anxiety features, which were found to constitute a separate factor [Eaton et al., 1989]. The checklist we used is not sensitive to anxiety symptoms and thus we did not detect any, though motor anxiety disturbances may be represented in the "excitement" factor. For analysis of factor scores we separated "major depression with psychotic features" from "major depression" as it constitutes a separate entity in the "delusion" factor. The "depression" factor scored very high on depressives, but psychotic disorder NOS and schizophreniform disorder also had some subjects scoring high. These two last diagnoses contain a great variety of psychopathologic manifestations, and many subjects presenting fringe schizophrenic or depressive disorders are in this category, a finding that may explain the score profile.

Negative schizophrenic symptoms are sometimes similar to depressive manifestations and might be scored as depressive symptoms. "Loss of pleasure" and "slowed activity" are examples of items present in both depressive states and schizophrenia. This may constitute a potential specificity bias, but the low presence of schizophrenic negative symptoms in depressed patients [Pogue-Geile and Harrow, 1984], and our finding of low scores of the "depression" factor in schizophrenics, indicate a low degree of overlap between depressive and schizophrenic negative symptoms.

The disorganization factor is very similar to that found by other researchers [Liddle and Barnes, 1990; Andreasen et al., 1995; Bilder et al., 1985]. A difference is that our disorganization factor also comprises the item "blunted affect," commonly included in the negative factor. However, it should be noted that these two factors often present a substantial degree of correlation, and actually the negative and disorganized factors of Andreasen et al. [1995], Peralta et al. [1994], and Maziade et al. [1995] partially overlap.

The last factor, delusional, is identical to the positive factor as defined by previous factorializations [Bilder et al., 1985; Liddle and Barnes, 1990; Perlata et al., 1994; Andreasen et al., 1995], and it is composed of delusions and hallucinations. The "disorganization" factor profile across diagnoses is quite interesting: the intermediate scores of psychotic disorder NOS and schizophreniform disorder further confirm the heterogeneous composition of these categories. Moreover, the high scores of some bipolar subjects reveal that during the most severe forms of mania the phenomenology largely overlaps with that of schizophrenic disorders. Regarding the "delusion" factor, subjects with schizophrenia, delusional, and schizophreniform disorders

produced the higher scores. An intermediate score was shown by depressive patients with psychotic features and by psychotic disorder NOS. This last point is noteworthy as it shows that, on average, psychotic depression and psychosis NOS have the same level of delusional symptoms. This finding strengthens the hypothesis that "major depression with psychotic features" has a separate and quite different phenomenological presentation from "major depression without psychotic features," and that the diagnosis of psychotic disorder NOS is highly heterogeneous, an impression based also on its scoring profile in the other factors. Thus, "delusion" should be viewed as a nonspecific factor present in more diagnoses.

The latent-roots approach allows for a more detailed definition of phenotypes in basic psychopathologic areas, and this can be accomplished by assigning a factor score to every patient and then using these scores for comparisons [Rutter, 1994; Cloninger, 1994; Lieberman, 1995]. Each subject is indeed defined by a continuous measure on any of the factors, e.g., a patient showing mainly depressive and delusional features may score like this (subject number 468): factor 1, "excitement" = −0.550; factor 2, "depression" = 2.365; factor 3, "disorganization" = −1.61; factor 4, "delusion" = 0.942.

The derivation of factor scores is an important advantage of the factor analytic technique over other methods such as cluster analysis or multidimensional scaling. Cluster analysis is basically an iterative procedure for grouping objects in the n-dimensional space composed by the variables considered, and has been used by some researchers for schizophrenia and major depression [Minas et al., 1992; Maes et al., 1994]. Grouped objects may be subjects or variables. Grouping subjects is a very powerful method to detect homogenous subgroups of subjects, particularly when a large number of variables is considered. Grouping variables has the same objective as factor analysis, i.e., to find highly correlated groups of variables, but it has been argued that the results obtained are not stable since they are strongly dependent on the algorithm used (hierarchical single linkage, hierarchical average linkage, k-means) [Milligan and Glen, 1981].

The derivation of factor scores allows several possible consequences. The immediate and intuitive use of our findings would be to apply quantitative measures to categorical diagnoses. Such a measure could then be used as a criterion to separate subgroups of patients within a discrete diagnosis, based on prevalence of a symptomatic pattern or severity of clinical condition. This is possible considering that higher scores correspond to a greater severity of disease or to longer duration of symptoms. In addition to those applications, however, there are other possible and less obvious uses of our findings.

The shift from an all-or-none trait to a semiquantitative phenotype could allow an investigator to explore the genetic determinants of mood and schizophrenic disorders as quantitative trait loci (QTLs) or threshold models. Actually, at least in psychiatric genetics, thresholds or liability models have mostly relied on formal genetic techniques, since the existing approaches were essentially based on likelihood theories of segregation analyses, with the exception of the logistic approach for quantitative phenotypes originally proposed by Bonney [1984]. Compared to this approach, the use of factor-derived scores in determining phenotype would allow for a more direct possibility of testing for correlation between phenotype and a given gene.

A limitation of the present study is its cross-sectional nature, with major focus on a "prevalence" design compared to a follow-up strategy. However, to correct for this bias, we used a lifetime perspective to detect "ever-occurred" symptoms. Other limitations could be ascribed to using mood and schizophrenic spectrum disorders pooled together. An admixture of two distinct, discrete diagnostic categories could lower the power of detecting subsyndrome factors, but this is also the positive peculiarity of the study, i.e., to detect structures common to these psychotic disorders. Additionally, a common limitation of all factor analytic studies is that the variance explained by the factors is always a fraction of the whole variance of the data. In our case, the variance explained is 54.63%, and even if the fitting indices show a very good fit of the model, a portion of variance remains unexplained and is considered measurement error. A part of this variance is certainly due to measurement error, but there is the possibility that another part of variance, i.e., some characteristics of a subject, is not included by the model and thus ignored. As already pointed out in several previous studies, the use of drugs could bias our ability to detect the naive symptomatology of a given psychosis, but it has been shown that this should not constitute a major obstacle in detecting valid latent structures [Liddle and Barnes, 1990; Peralta et al., 1994].

In conclusion, given the procedures that we have applied throughout the analysis (EFA, extended EFA, and CFA), the four factors extracted resulted in: 1) Having a strong construct and content validity; 2) Having been replicated in an independent sample; and 3) Having factor-score distributions that properly reconstructed diagnostic categories. This last point is interesting, since both diagnoses and factors point to the same underlying structures: the "excitement" factor mostly overlaps with the diagnosis of manic episode, the "depression," factor with "major depression," and the last two factors ("disorganized" and "delusional") with "schizophrenia disorders."

The originality of this study is that the search for latent structures was carried out in a very large sample, and that it was composed of both schizophrenic and mood disorders. The use of factor scores derived from these factors could represent a very sensitive phenotypic measure for genetic studies, in addition to traditional categorical diagnoses: practical application of these semiquantitative traits ranges from traditional linkage to nonparametric, i.e., sib-pair association studies.

REFERENCES

American Psychiatric Association (1987): "Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition." Washington, DC: American Psychiatric Association, pp 27–365.

Andreasen NC, Olsen S (1982): Negative and positive schizophrenia: Definition and validation. Arch Gen Psychiatry 39:789–794.

Andreasen NC, Arndt S, Alliger R, Miller D, Flaum M (1995): Symptoms of schizophrenia. Methods, meanings and mechanisms. Arch Gen Psychiatry 52:341–351.

Bilder RM, Mukergee S, Rieder RO, Pandurangi AK (1985): Symptomatic and neuropsychological components of defect state. Schizophr Bull 11:409–419.

Bollen KA (1989): "Structural Equations With Latent Variables." New York: John Wiley & Sons, pp 226–317.

Bonney GE (1984): On the statistical determination of major gene mechanisms in continuous human traits: Regressive models. Am J Med Genet 18:731–749.

Cloninger CR (1994): Turning point in the design of linkage studies of schizophrenia. Am J Med Genet 54:83–92.

Crow TJ (1980): Molecular pathology of schizophrenia: More than one disease process? Br Med J [Clin Res] 280:66–68.

Dixon WJ (1990): "BMDP Statistical Software Manual." Berkeley: University of California Press, pp 311–337.

Eaton WW, Dryman A, Sorenson A, McCutcheon A (1989): DSM-III major depressive disorder in the community. A latent class analysis of data from the NIMH epidemiologic catchment area programme. Br J Psychiatry 155:48–54.

Eaves LJ, Silber JL, Hewitt JK, Rutter M, Meyer JM, Neale MC, Pickles A (1993): Analyzing twin resemblance in multisymptom data: Genetic applications of a latent class model for symptoms of conduct disorder in juvenile boys. Behav Genet 23:5–19.

Endicott J, Spitzer RL (1978): A diagnostic interview: The schedule for affective disorders and schizophrenia. Arch Gen Psychiatry 35: 773–782.

Faraone SV, Tsuang MT (1994): Measuring diagnostic accuracy in the absence of a "gold standard." Am J Psychiatry 151:650–657.

Farmer AE, Williams J, Jones I (1994): Phenotypic definitions of psychotic illness for molecular genetic research. Am J Med Genet 54: 365–371.

Joreskog KG, Sorbom D (1989): "LISREL 7—A Guide to the Program and Applications," 2nd ed. Chicago: SPSS Publications, pp 2–25.

Joreskog KG, Sorbom D (1993): "LISREL 8—Users Guide." Chicago: SSI Publications, pp 134–137.

Kendler KS, Heat AC, Martin NG, Eaves LJ (1987): Symptoms of anxiety and symptoms of depression. Arch Gen Psychiatry 44: 451–457.

Klimidis S, Stuart GW, Minas IH, Copolov DL, Singh BS (1993): Positive and negative symptoms in the psychoses. Re-analysis of published SAPS and SANS global ratings. Schizophr Res 9:11–18.

Liddle PF, Barnes TRE (1990): Syndromes of chronic schizophrenia. Br J Psychiatry 157:558–561.

Lieberman JA (1995): Signs and symptoms. What can they tell us about the clinical course and pathophysiologic processes of schizophrenia? Arch Gen Psychiatry 52:361–362.

Maes M, Meltzer HY, Cosyns P, Schotte C (1994): Evidence for the existence of major depression with and without anxiety features. Psychopathology 27:1–13.

Maziade M, Roy MA, Martinez M, Cliche D, Fournier JP, Garneau Y, Nicole L, Montgrain N, Dion C, Ponton AM, Potvin A, Lavallée JC, Pirès A, Bouchard S, Boutin P, Brisebois F, Merette C (1995): Negative, psychoticism, and disorganized dimensions in patients with familial schizophrenia or bipolar disorder: Continuity and discontinuity between the major psychoses. Am J Psychiatry 152:1458–1463.

McGuffin P, Farmer A, Harvey I (1991): A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. Arch Gen Psychiatry 48: 764–770.

Milligan GW (1981): A Monte Carlo study of thirty internal criterion measures for cluster analysis. Psychometrika 46:187–199.

Minas IH, Stuart GW, Klimidis S, Jackson HJ, Singh BS, Copolov DL (1992): Positive and negative symptoms in the psychoses: Multidimensional scaling of SAPS and SANS items. Schizophr Res 8: 143–156.

Mortimer AM, Lund CE, McKenna PJ (1990): The positive-negative dichotomy in schizophrenia. Br J Psychiatry 157:41–49.

Parker G, Hadzi Pavlovic D, Brodaty H, Boyce P, Mitchell P, Wilhelm K, Hickie I, Eyers K (1993): Psychomotor disturbance in depression: Defining the constructs. J Affective Disord 27:255–265.

Peralta V, Cuesta M, de Leon J (1994): An empirical analysis of latent structures underlying schizophrenic symptoms: A four-syndrome model. Biol Psychiatry 36:726–736.

Pogue-Geile MF, Harrow M (1984): Negative and positive symptoms in schizophrenia and depression: A follow up. Schizophr Bull 10: 371–387.

Reyment RA, Joreskog KG (1993): "Applied Factor Analysis in the Natural Sciences." Cambridge: Cambridge University Press.

Robins LN, Helzer J, Croughan J (1981): The NIMH diagnostic interview schedule: Its history, characteristics, and validity. Arch Gen Psychiatry 39:381–389.

Rutter M (1994): Psychiatric genetics: Research challenges and pathways forward. Am J Med Genet 54:185–198.

Tsuang MT (1993): Genotypes, phenotypes, and the brain. A search for connections in schizophrenia. Br J Psychiatry 163:299–307.

Tsuang MT, Faraone SV (1990): "The Genetics of Mood Disorders." Baltimore: The Johns Hopkins University Press, pp 1–30.

Von Knorring L, Lindstrom E (1995): Principal components and further possibilities with the PANSS. Acta Psychiatr Scand [Suppl] 91:5–10.

Young MA, Abrams R, Taylor MA, Meltzer HY (1983): Establishing diagnostic criteria for mania. J Nerv Ment Dis 171:676–682.